CAM

CAM

Centre for Applied Microeconometrics

Institute of Economics
University of Copenhagen

http://www.econ.ku.dk/CAM/

Nonlinear Models with Panel Data

Bo Honoré

2002-02

The activities of CAM are financed by a grant from
The Danish National Research Foundation

Nonlinear Models with Panel Data

Bo E. Honoré*
Department of Economics, Princeton University

February, 2002

1 Introduction

There is a natural grouping of the observations in many economic data sets. For example, a data set might contain information about an individual or a firm over a number of time periods. In this case one can think of the observations for a given individual or firm as a group. This kind of data set is often referred to as a panel data set, but some of the tools that have been developed for this case are also applicable in other situations where the observations are grouped. It is therefore natural to use the term panel data for any situation in which there is a natural grouping of the data. For example, a data set might contain information on households, in which case one might think of the household living in the same narrowly defined geographical area as a group. In a data set of individuals, one might think of individuals that belong the same family as a group. There are at least two reasons why such a grouping is interesting, — and important, even if it is not interesting. The first reason is that with grouped data, one would expect the errors in an econometric model to be related across observations in the same group. This raises econometric issues that must be dealt with even if this relationship is not the object of interest. This correlation within a group is often modeled by allowing for a group–effect which has the interpretation as an unobserved group–specific explanatory variable. The second motivation for studying grouped data is that the relationship between the observations in a given group might be interesting in itself. For example, if the group consists of individuals over time, then it is sometimes of interest to know

*The research behind this paper was supported by the National Science Foundation, the Gregory C. Chow Econometric Research Program at Princeton University, and Danish National Research Foundation (through CAM at the University of Copenhagen). The author thanks Marina Sallustro for helpful suggestions.

how (and whether) the dependent variable in one period affects its future values.

This paper will discuss some issues related to panel data estimation of standard nonlinear econometric models. The paper is not intended as a survey, but rather as an introduction to a particular subset of the literature. Although some of the tools described here apply to any kind of grouped data, we will use the terminology that one would use if the data consists of n individuals observed over T time periods. The generic model that will be discussed has the form

$$y_{it} = g\left(x_{it}, \varepsilon_{it}, \alpha_i; \theta\right) \tag{1}$$

where y_{it} is the dependent variable of interest and x_{it} is a vector of explanatory variables for individual i in time period t. We will use y_i and x_i (without the subscript t) to denote a vector of all the dependent (and explanatory) variables for individual i. α_i is a time–invariant, individual specific effect, which can be interpreted as an unobserved explanatory variable. α_i is sometimes referred to as unobserved heterogeneity. θ is the vector of parameters to be estimated, and almost all of the results that will be discussed for estimators of nonlinear panel data models, are justified by asymptotics rather than finite sample arguments. The asymptotic arguments assume that n is large with small (fixed) T, which corresponds to the situation in which many individuals are observed in few time periods.

We will discuss a number of different nonlinear models in this paper, but the leading specific example to be considered is the discrete choice model

$$y_{it} = \begin{cases} 1 & \text{if} \quad x_{it}\beta + \alpha_i + \varepsilon_{it} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

If ε_{it} is normally or logistically distributed and independent of (x_{it}, α_i) then this is a panel data version of the familiar probit and logit models. Note that one cannot identify the scale of β and ε in (2). When discussing estimation of (2), we will therefore assume that some scale normalization is imposed.

Distinguishing between the assumptions that are made with regard to the relationship between current dependent variables and future explanatory variables is crucial in the analysis of panel data. This is most easily seen by considering the case where (1) is linear. In that case, first–differencing will eliminate the α_i,

$$(y_{it} - y_{it-1}) = (x_{it} - x_{it-1})'\beta + (\varepsilon_{it} - \varepsilon_{it-1})\,. \tag{3}$$

If the error in time period t, ε_{it}, is uncorrelated with past, current and future values of the explanatory variable, then one can estimate β by applying ordinary least squares to (3). On the other hand, if the error is only uncorrelated with past and current values of the explanatory variables, then $(\varepsilon_{it} - \varepsilon_{it-1})$ will be correlated with $(x_{it} - x_{it-1})$ and the ordinary least squares estimator will be inconsistent.[1] In this paper we will say that the explanatory variables are *strictly exogenous* when assumptions are made on the distribution of the errors conditional on past, current and future values of the explanatory variables. We will say that explanatory variables are *predetermined* when assumptions are made on the distribution of the errors conditional only on past and current values of the explanatory variable.

2 Static Models

It is useful to first consider models in which the explanatory variables are strictly exogenous. This assumption is unrealistic in most economic contexts if one interprets t as time. The reason is that in that case, one typically wants to think of y_{it} as the outcome of some agent's optimization problem, and since the agent (presumable) observes y_{it} before solving the optimization problem that leads to y_{it+1}, one expects y_{it} to be an explanatory variable in the equation for y_{it+1}, which would rule out strict exogeneity. Moreover, in many economic examples, x_{it} is itself a choice variable and one may want to allow for the possibility that the agent chooses x_{it} on the basis of past values of y_{it}.[2] On the other hand, as pointed out in the introduction, the panel structure does not necessarily have to be due to the agents being observed over time. If, for example, t in (1) refers to a family in a village, then it might be reasonable to assume that the y_{it}s depend on each other only through the explanatory variables x_{it} and through some village specific unobserved characteristic, α_i. In other words, in that case it might be reasonable to make assumptions on the errors conditional on the explanatory variables for all the units in a particular group. Specifically, in (2) one might assume that the errors ε_{it} are independent and all logistically or normally distributed conditional

[1] In this case one, can estimate β by an instrumental variables approach that uses past values of the explanatory variables as instruments. See e.g., Arellano and Bond (1988).

[2] An often cited example of this is models of labor supply. In those models, the presence of small children in the household is often used as an explanatory variable. In this case, it might be reasonable to assume that shocks to the labor supply are unrelated to current and past values of the number of small children, but it is much more difficult to justify an assumption that it is also unrelated to future values, because that would imply that shocks to the labor supply have no effect on the current and future decisions regarding child–bearing. See, e.g., Browning (1992).

on[3] (α_i, x_i). In that case (2) becomes a logit or probit model with group–specific constants.

Assume for the moment that one is willing to parameterize the distribution of ε_i given (α_i, x_i) in models like (1) and (2). There are then essentially two approaches that have been taken to deal with the group–specific effect, α_i. They are, perhaps somewhat misleadingly, referred to as the *random effects* and the *fixed effects* approach.

In a random effects approach, one parameterizes the distribution of α_i conditional on x_i. This makes the model fully parametric and it can, in principle, be estimated by maximum likelihood. For example, if one assumes that ε_i in (2) is independent of (α_i, x_i) with a multivariate normal distribution, and that α_i is normally distributed and independent of x_i, then one can express the probability of any sequence of ys in terms of a multivariate normal distribution. Of course, because of numerical issues related to how quick and accurate calculation of the multivariate normal CDF, it can be very difficult to actually use these probabilities to estimate the model by maximum likelihood. The large literature on simulation–based estimation[4] is therefore very much relevant, although the developments in that literature are not specific to panel data. Because the random effects approach makes the model fully parametric, it is also conceptually straightforward to approach the estimation of its parameters from a Baysian point of view.

In a pure random effects model, one can also ignore the panel structure and estimate the model by a pseudo–maximum likelihood method that ignores the panel structure altogether. Once the distributions of ε_i and α_i have been specified, one can obtain the distribution of y_{it} given x_{it} in (1), and one can then estimate the parameters of interest, θ, by ignoring the panel structure and treating the data as one large cross section. Under suitable regularity assumptions, this will lead to a consistent and asymptotically normal estimator, although one will have to correct the standard errors for the fact that the observations are not independent.[5] Consider for example a random effects probit model. If ε_i is composed of *i.i.d.* normally distributed random variables, then the marginal distribution for all the ys are as in a probit model, and all the marginal choice probabilities

[3]Recall that x_i is defined to be $(x_{i1}, ..., x_{iT_i})$

[4]See e.g. Hajivassiliou and Ruud (1994)

[5]Estimating the panel model by considering only the marginal distribution of y_{it} (and not the joint) will typically not lead to estimates of the parameters of the distribution of the errors, ε_{it} and α_i. One possibility is to proceed in a two–step manner where one first estimates θ as described, and then estimates the parameters of the distributions of ε_{it} and α_i using features of the joint distribution of the vector y_i.

(and β) can be estimated by treating the data as one large cross section.

In a fixed effects approach, one attempts to find ways to estimate θ in (1) making only minimal assumptions on α_i. This is inspired by the linear panel data model, in which one can difference away the group specific effects as in (3). One can also motivate it by noting that if one had only few individuals observed over many time periods, then one could justify treating the α_is as parameters to be estimated by assuming that the relevant asymptotic arguments let T_i increase with fixed n. One could then proceed parametrically and estimate all the parameters by maximum likelihood (or some other convenient method). This would require no assumptions on the distribution of α_i. However, if one is in a situation where the number of individuals is large and the number of time periods small, then it seems more appealing to consider asymptotics with large n and fixed T_i. Treating the α_i's as parameters to be estimated in that case implies that the number of parameters will be increasing as the sample size increases. This problem, which is called the incidental parameters problem, will typically, but not always, lead to inconsistent estimation of all the parameters of the model (see Neyman and Scott (1948)). The contribution of the literature on estimation of nonlinear fixed effects panel data models has been to develop alternative estimation procedures for estimating θ in (1) without making assumptions on the distribution of the α_is. The general idea is that although the model does not have features that are linear in the $\alpha_i s$ (so the α_is cannot be differenced away), it is nonetheless sometimes possible to find features of the model that do not depend on α_i. These are the features that will be used to construct estimators of the models. Unfortunately, the features of the model that do not depend on α_i tend to be different for different functional forms for g in (1). The resulting estimation procedures are therefore different for different models, and one ends up estimating, say, a logit model in a way that is fundamentally different from the way one would estimate, say, a censored regression model. This is somewhat unsatisfactory and fundamentally different from the random effects model in which one can use one approach, such as maximum likelihood, to estimate a number of different models.

To see how one can estimate a nonlinear panel data model without any assumptions of the distribution of the α_is, consider (2) with independent logistic ε_{it}s and for simplicity assume that $T = 2$. In that case,

$$P\left(y_{i1} = d_1, y_{i2} = d_2 \middle| x_{i1}, x_{i2}, \alpha_i\right) = \frac{\exp\left(x_{i1}\beta + \alpha_i\right)^{d_1}}{1 + \exp\left(x_{i1}\beta + \alpha_i\right)} \frac{\exp\left(x_{i2}\beta + \alpha_i\right)^{d_2}}{1 + \exp\left(x_{i2}\beta + \alpha_i\right)} \tag{4}$$

and the feature of the model that do not depend on α_i is

$$P\left(y_{i1}=d \mid x_{i1}, x_{i2}, \alpha_i, y_{i1}+y_{i2}=1\right)=\frac{\exp\left(\left(x_{i1}-x_{i2}\right)\beta\right)^{d}}{1+\exp\left(\left(x_{i1}-x_{i2}\right)\beta\right)}$$

In other words, for the individuals for whom y changes, the probability that it changes from 1 to 0, as opposed to changing from 0 to 1, is a logit with explanatory variables $(x_{i1}-x_{i2})$. Since this probability does not depend on α_i, one can estimate β without making assumptions on α_i by considering only the individuals for whom $y_{i1}+y_{i2}=1$, and then estimating a logit model for the event $y_{i1}=1$. It is intuitively appealing that the individuals who do not switch, are not used to estimate β, since for any value of β, those individuals can be rationalized either by extremely large or by extremely small values of α_i. However, a result due to Hahn (2001) suggests that the estimator of β based on maximum likelihood estimation conditional on $y_{i1}+y_{i2}=1$ may not be asymptotically efficient.

The fixed effects logit model in (4) also illustrates a fundamental difficulty estimating nonlinear models. Knowing β in (2) allows one to judge the relative importance of different time–varying explanatory variables and it also allows one to calculate the effect of x_{it} on the probability that $y_{it}=1$ conditional on a particular value for α_i. However, it does not allow one to calculate the average effect of x_{it} on the probability that $y_{it}=1$ across the distribution of α_i in the population. See Wooldridge (2000) for a lenghty discussion of this. This phenomenon is not specific to panel data models, but rather a general feature of many nonlinear semiparametric models. Consider for example the cross sectional semiparametric discrete choice model

$$y_i=1\left\{x_i\beta+\varepsilon_i\geq 0\right\}$$

where ε_i is independent of x_i. Many papers in econometrics have considered estimation of β in this model,[6] but knowledge of β is not sufficient for one to calculate the effect of x_i on the probability that $y_i=1$. On the other hand, knowing β allows one to judge the relative importance of the components of x_i and to determine which components have no effect. Also, if the model being investigated is derived from some structural economic model, then the parameter might be of independent interest.

One appealing aspect of the fixed effects approach is that α_i is allowed to depend on x_i in an arbitrary manner. A number of authors (see for example Chamberlain (1984), Newey (1994)

[6] See for example Powell (1994) for a discussion of this and references to this literature.

and Chen (1998)) have tried to accomplish this in a random effects approach by parameterizing the distribution of α_i as a function of x_i. This is fairly innocent in a linear model, since one can interpret the coefficient as the parameters in linear projections, and consequently there is a sense in which the model is always correctly specified. Unfortunately, this is less true for nonlinear models. Consider for example a probit version of (2). In that case it would be natural to assume that

$$\alpha_i | \left(x_{i1}, ..., x_{iT_i}\right) \sim N\left(\sum_{t=1}^{T_i} x_{it}\gamma_t, \sigma_\alpha^2\right) \tag{5}$$

where the parameters γ_t and σ_α^2 might depend on T_i. The marginal distribution of y_{it} would be from a probit model and it would be easy to estimate the parameters of interest (subject to the normalizations necessary for identification). Unfortunately, it is difficult to justify (5) because it one would presumably want it to hold no matter what T_i happens to be in a particular sample. But, as a general statement, this places strong assumptions on the distribution of the explanatory variables. In particular, the law of iterated expectations implies that if (5) holds in time periods T and $T+1$ with γ_t given by γ_t and $\widetilde{\gamma}_t$, respectively, then $E\left[\sum_{t=1}^{T+1} x_{it}\widetilde{\gamma}_t \middle| x_{i1}, ..., x_{iT}\right] = \sum_{t=1}^{T} x_{it}\gamma_t$ or $E\left[x_{iT+1}| x_{i1}, ..., x_{iT}\right] = \left(\sum_{t=1}^{T} x_{it}\left(\gamma_t - \widetilde{\gamma}_t\right)/\widetilde{\gamma}_{T+1}\right)$. In other words, the mean of x_{iT+1}, given its past values, is not only linear, but the coefficients are linked to the ones one would get by estimating a probit model for the distribution of y_{it} given the explanatory variables in various periods. (5) has many other implications of the same spirit. Of course, even if one was not willing to accept these restrictions on the distribution of the regressors and on the parameters, one might still be willing to proceed from (5) on the basis that it might be a useful approximation that captures the possibility that α_i is related to x_i, and allows one to estimate all the parameters needed for calculating the effect of x_i on the probability that $y_{it} = 1$.

A recent paper by Altonji and Matzkin (2001) takes a more nonparametric approach to estimation of (1). In order to simplify the exposition, assume that ε_i is independent of x_i. The basic idea is that the value of x_{it} affects the distribution of y_{it} both directly and indirectly through its effect on α_i. On the other hand, it will only affect the distribution of y_{is} indirectly through its effect on α_i. With a little additional structure, one can then use the joint distributions of y_{is} and x_{it} and y_{it} and x_{it}, respectively, to estimate the direct effect of x_{it} on the distribution of y_{it}, holding α_i fixed. While most of the results in Altonji and Matzkin (2001) are nonparametric in the sense that they are concerned with estimating the effect on the distribution of y_{it}, it is in principle possible to turn these into estimates of the effect of x_{it} on (say) the mean of y_{it}.

2.1 Details on some specific nonlinear models

As mentioned above, it is possible to estimate the coefficients on x_{it} in a panel data logit model with two time periods by considering the conditional distribution of y_i given $y_{i1} + y_{i2}$. This is a special case of a more general idea. Consider the model given in (1). A sufficient statistic, S_i, for α_i is defined to be a function[7] of the data such that the distribution of y_i conditional on (S_i, x_i, α_i) does not depend on α_i. In the two–period logit model, $y_{i1} + y_{i2}$ is such a function. If one has a sufficient statistics which furthermore has the property that the distribution of y_i conditional on (S_i, x_i, α_i) depends on θ, then one can estimate θ by maximum likelihood using the conditional distribution of the data, given the sufficient statistics. Andersen (1970) proved that the resulting estimator is consistent and asymptotically normal under appropriate regularity conditions.

Unfortunately, there are only few standard nonlinear econometric panel data models for which a sufficient statistic that has the appropriate properties exist. For example, the only sufficient statistic in the probit version of (2) is y_i, and it is therefore clear that one cannot make inference about β in by considering the distribution conditional on the sufficient statistic. A second limitation of the approach of conditioning on sufficient statistics is that it requires a parametric model for y_i conditional on (x_i, α_i). In this subsection we will therefore discuss some alternative approaches to estimation of nonlinear fixed effects panel data models. One can interpret these as generalizations of the conditional likelihood approach. Specifically, the general idea is to look for some feature of the data, whose distribution depends on θ, but not on α_i. That feature will then be used to estimate θ without making assumptions on α_i. In the conditional likelihood approach, the feature used to estimate θ, is the conditional distribution of y_i given (S_i, x_i, α_i). In the approaches discussed below, the features will be objects like moments and medians. Before discussing these, we will briefly review two cases in which the conditional likelihood approach does work.

Example 1 (Logit) *The sufficient statistic for a logit model with T_i observations for each individual is $S_i = \sum_{t=1}^{T_i} y_{it}$ and the conditional distribution of y_i given (S_i, x_i, α_i) is*

$$P\left(y_{i1}, ..., y_{iT_i} | \sum_{t=1}^{T_i} y_{it}, x_i, \alpha_i\right) = \frac{\exp\left(\sum_{t=1}^{T_i} y_{it} x_{it} \beta\right)}{\sum_{(d_1, ..., d_{T_i}) \in B} \exp\left(\sum_{t=1}^{T_i} d_t x_{it} \beta\right)}$$

Example 2 (Poisson Regression) *The panel data poisson regression model is given by*

$$y_{it} | x_i, \alpha_i \sim \text{po}\left(\exp(\alpha_i + x_{it}\beta)\right)$$

[7]The sufficient statistic, S_i, does not have to be a scalar.

For this model, the sufficient statistic is $\sum_{t=1}^{T_i} y_{it}$ *and the conditional distribution of* y_i *given* (S_i, x_i, α_i) *is*

$$P\left(y_{i1},...,y_{iT_i}|\sum_{t=1}^{T_i} y_{it}, x_{i1},...,x_{iT_i},\alpha_i\right) = \frac{\left(\sum_{t=1}^{T_i} y_{it}\right)!}{\prod_{t=1}^{T_i} y_{it}!}\prod_{t=1}^{T_i}\frac{\exp(x_{it}\beta)}{\sum_{s=1}^{T_i}\exp(x_{is}\beta)} \tag{6}$$

This follows from the fact that the conditional distribution of independent poisson random variables is multinomial.

Conditional likelihood estimation of the panel data poisson regression model was considered by Hausman, Hall, and Griliches (1984). Somewhat surprisingly, Blundell, Griffith, and Windmeijer (1997) and Lancaster (1997) have shown that treating all the fixed effects as parameters to be estimated, leads to the conditional maximum likelihood estimator based on (6). In other words, the incidental parameters problem does not lead to an inconsistent estimator in the panel data poisson regression model with strictly exogenous regressors.

The next three examplesa illustrate how some features of the model, other than the conditional likelihood, can depend on the parameter to be estimated, but not of α_i, and how this can be used to construct an estimator. For all of them, the basic idea is to compare two observations for a given individual. So to implement the ideas in practice, one will have to consider all pairs of time periods and then combine them in some way. The three examples are interesting in their own right, but the main reason for presenting them here, is to illustrate the close link between the literature on non–linear fixed effects panel data models and the estimation of nonlinear semiparametric cross sectional models.

Example 3 (Semiparametric Binary Choice) *Consider a semiparametric version of the panel data discrete choice model* (2)*, where, for two time periods* t *and* s*,* ε_{it} *and* ε_{is} *are identically distributed conditional on* $(\alpha_i, x_{it}, x_{is})$*. Manski (1987) observed that in that model*

$$\text{median}\left(y_{it} - y_{is}|\, \alpha_i, x_{it}, x_{is}, y_{it} \neq y_{is}\right) = \text{sign}\left((x_{it} - x_{is})\,\beta\right) \tag{7}$$

where the key observation is that the right-hand side does not depend on α_i *but that it does depend on* β*. One can therefore estimate* β *by minimizing*

$$\sum_{i=1}^{n}\left|(y_{it} - y_{is}) - \text{sign}\left((x_{it} - x_{is})\,b\right)\right|$$

which is equivalent to maximizing

$$\sum_{i=1}^{n}\text{sign}\left(y_{it} - y_{is}\right)\cdot\text{sign}\left((x_{it} - x_{is})\,b\right)$$

The resulting estimator is consistent but not root–n asymptotically normal.

To understand the insight behind Example 3, one must relate it to Manski's earlier work on the cross sectional binary choice model (Manski (1975) and Manski (1985))

$$y_i = \begin{cases} 1 & \text{if} \quad x_i\beta + \varepsilon_i \geq 0 \\ 0 & \text{otherwise} \end{cases}, \qquad \text{with median}\,(\varepsilon_i | \, x_i) = 0 \tag{8}$$

If $\varepsilon_i = x_i\beta$ with probability 0, then(8) implies that $2y_i - 1 = \text{sign}\,(x_i\beta + \varepsilon_i)$ (with probability 1). This implies that $\text{median}\,(2y_i - 1|\, x_i) = \text{sign}\,(x_i\beta)$, which in turn implies that the function $E\,[\text{sign}\,(2y_i - 1) \cdot \text{sign}\,(x_i b)]$ is maximized at $b = \beta$. It is therefore natural to estimate β by maximizing $\sum_{i=1}^{n} \text{sign}\,(2y_i - 1) \cdot \text{sign}\,(x_i b)$.

The key observation in Manski (1987) is that in the panel data version of the model, the behavior of $y_{it} - y_{is}$ conditional on $y_{it} \neq y_{is}$ is similar to the behavior of $2y_i - 1$ in the cross sectional version given in (8). With that observation, one can then proceed just as one would in the cross sectional case, and the properties of the resulting estimator are essentially identical to the properties of the maximum score estimator proposed in Manski (1975).

Example 4 (Censored Regression) *Honoré (1992)considered the censored regression model*

$$y_{it} = \max\{0, x_{it}\beta + \alpha_i + \varepsilon_{it}\} \tag{9}$$

and showed that if $(\varepsilon_{it}, \varepsilon_{is})$ is distributed like $(\varepsilon_{is}, \varepsilon_{it})$ conditional on $(x_{it}, x_{is}, \alpha_i)$ then β is the unique minimizer (as a function of b) of the function

$$\begin{aligned} &E\Big[(\max\{y_{it}, \triangle x_i b\} - \max\{y_{is}, -\triangle x_i b\} - \triangle x_i b)^2 \\ &+2 \cdot 1\{y_{it} < \triangle x_i b\}(\triangle x_i b - y_{it})y_{is} + 2 \cdot 1\{y_{is} < -\triangle x_i b\}(-\triangle x_i b - y_{is})y_{it}\,\Big] \end{aligned} \tag{10}$$

where $\triangle x_i = x_{it} - x_{is}$. This suggests estimating β by

$$\begin{aligned} \widehat{\beta} \;=\; &\arg\min_b \sum_{i=1}^{n} \Big((\max\{y_{it}, \triangle x_i b\} - \max\{y_{is}, -\triangle x_i b\} - \triangle x_i b)^2 \\ +2 \cdot 1\{y_{it} \;<\; &\triangle x_i b\}(\triangle x_i b - y_{it})y_{is} + 2 \cdot 1\{y_{is} < -\triangle x_i b\}(-\triangle x_i b - y_{is})y_{it} \Big). \end{aligned} \tag{11}$$

The minimization problem (10) is convex and has as first order condition

$$0 = E\,[((\max\{y_{it}, \triangle x_i b\} - \max\{0, \triangle x_i b\}) - (\max\{y_{is}, -\triangle x_i b\} - \max\{0, -\triangle x_i b\}))\,\triangle x_i]$$

at $b = \beta$ the right hand side equals

$$E\left[\left(\max\left\{\alpha_i + \varepsilon_{it}, -x_{it}\beta, -x_{is}\beta\right\} - \max\left\{\alpha_i + \varepsilon_{is}, -x_{is}\beta, -x_{it}\beta\right\}\right)\triangle x_i\right]$$

If ε_{it} and ε_{is} are identically distributed conditional on $(\alpha_i, x_{it}, x_{is})$, then this is clearly 0. That is the reason why (10) is minimized at $b = \beta$[8]. As discussed in Honoré and Kyriazidou (2000a) and Arellano and Honoré (2001), there are many other estimator of β in (9). Moreover, the estimator in (11) above is consistent and asymptotically normal under Manski (1987)'s assumption that ε_{it} and ε_{is} are identically distributed conditional on $(\alpha_i, x_{it}, x_{is})$. This is weaker than the assumption made in Honoré (1992). It is trivial to modify the model in such a way that the same estimation strategy applies to the models $y_{it} = \max\{c_i, x_{it}\beta + \alpha_i + \varepsilon_{it}\}$ or $y_{it} = \min\{c_i, x_{it}\beta + \alpha_i + \varepsilon_{it}\}$ where c_i is observed.

The censored regression model can be motivated in many ways. Perhaps the cleanest is the case where some relationship of interest is assumed to be as in the linear panel data model

$$y_{it}^* = x_{it}\beta + \alpha_i + \varepsilon_{it}$$

but where the dependent variable of interest, y_{it}^*, is subject to top coding, so the observed variable is $\min(c_t, y_{it}^*)$, where c_t is the value above which y_{it}^* is top coded. In that case, the parameter β is the marginal effect of x on the variable of interest y^*. So while the model is nonlinear, and β, therefore, is not the marginal effect of x on the observed y, it is the marginal effect of x on the dependent variable of interest.

Example 5 (Sample Selection) *Kyriazidou (1997) studied the following panel data version of the "standard" sample selection model*

$$\begin{aligned} y_{1it}^* &= x_{1it}\beta_1 + \alpha_{1i} + \varepsilon_{1it} \\ y_{2it}^* &= x_{2it}\beta_2 + \alpha_{2i} + \varepsilon_{2it} \end{aligned}$$

where we observe:

$$y_{1it} = 1\{y_{1it}^* > 0\} \tag{12}$$

$$y_{2it} = \begin{cases} y_{2it}^* & \textit{if } y_{1it} = 1 \\ 0 & \textit{otherwise} \end{cases} \tag{13}$$

[8] Honoré (1992) also gives a graphical motivation for (10). This can be seen as a two–dimensional version of the graphical motivation of the estimator of the cross sectional censored regression model proposed in Powell (1986).

and it is assumed that the errors $(\varepsilon_{1it}, \varepsilon_{2it})$ *are independent*[9] *and identically distributed and independent of* (α_i, x_i). *Since the model for* y_{1it} *is a discrete choice model like the one discussed in Examples 1 and 3, the new challenge is to estimate* β_2. *Kyriazidou (1997) showed that if* $K(\cdot)$ *is a kernel, and* h_n *is a sequence of numbers that converges to 0 at an appropriate rate as* n *increases, then*

$$\begin{aligned}\hat{\beta}_2 &= \left[\sum_{i=1}^{n} (x_{2it} - x_{2is})' (x_{2it} - x_{2is}) K\left(\frac{(x_{1it} - x_{1is})\hat{\beta}_1}{h_n}\right) y_{1it} y_{1is}\right]^{-1} \\ &\quad \times \left[\sum_{i=1}^{n} (x_{2it} - x_{2is})' (y_{2it} - y_{2is}) K\left(\frac{(x_{1it} - x_{1is})\hat{\beta}_1}{h_n}\right) y_{1it} y_{1is}\right]\end{aligned} \qquad (14)$$

is consistent and asymptotically normal under appropriate regularity condition. However, the rate of convergence of $\hat{\beta}_2$ *is slower than* $\sqrt{n}$.

The estimator in (14) can be understood by observing that

$$\begin{aligned} E\left[y_{2it} \middle| y_{1it} = 1, x_{1i}, x_{2i}, \alpha_{1i}, \alpha_{2i}\right] &= x_{2it}\beta_2 + \alpha_{2i} \\ &\quad + E\left[\varepsilon_{2it} \middle| \varepsilon_{1it} > -x_{1it}\beta_1 - \alpha_{1i}, x_{1i}, x_{2i}, \alpha_{1i}, \alpha_{2i}\right] \end{aligned}$$

where the last terms is a sample selection term similar to the one in Heckman (1976) and Heckman (1979). If $(\varepsilon_{1it}, \varepsilon_{2it})$ is distributed like $(\varepsilon_{1is}, \varepsilon_{2is})$ (conditional on $x_{1i}, x_{2i}, \alpha_{1i}, \alpha_{2i}$) then the last term will be the same for an individual who happens to have $x_{1it}\beta_1 = x_{1is}\beta_1$, and for such an individual, the transformation $y_{2it} - y_{2is}$ will eliminate the fixed effect, α_{2i}, as well as the sample selection term. The estimator $\hat{\beta}_2$ uses all observed differences of the form $y_{2it} - y_{2is}$, but the term $K\left(\frac{(x_{1it}-x_{1is})\hat{\beta}_1}{h_n}\right)$ guarantees that in the limit, as n increases, only terms for which $x_{1it}\beta_1 = x_{1is}\beta_1$ will get any weight.

The estimator proposed by Kyriazidou (1997) is closely related to an approach proposed by Powell (1987)[10] for the cross sectional sample selection model. Rather than considering pairs of time periods for a given individual in a panel, Powell considered all pairs of individuals in a cross section. This results in a $\sqrt{n}$–consistent estimator, essentially because the effective number of terms used to define Powell's estimator is of order n^2h (as opposed to nh in Kyriazidou (1997)).

[9] This assumption is stronger than necessary. See Kyriazidou (1997) or Honoré and Kyriazidou (2000a) for details.

[10] See also Ahn and Powell (1993)

3 Dynamic Models

When the second dimension in a panel is time, it is often natural to assume that one of the explanatory variables in (1) is a past value of the dependent variable. In that case, it does not make sense to assume that the explanatory variables are strictly exogenous as was done in the previous section.[11] The same is true when the explanatory variables x_{it} are allowed to depend on past values of the dependent variable. In this section, we will briefly discuss this problem and some proposed solutions in the case where the model has a set of strictly exogenous explanatory variables, x_{it}, as well as a lagged dependent explanatory variable.

As mentioned in Section 2, there are essentially two approaches that one can take to estimate nonlinear panel data models. A random effects in which one models the distribution of the individual specific effect, and a fixed effects approach in which the distribution of the individual specific effect is left completely unspecified. The trade-off between the two approaches is the same for dynamic models as it was for static models. In particular, for nonlinear models like (1) it is often true that knowing θ does not allow one to calculate marginal effects of interest. In the other hand, the consistency of the estimators of a random effects model usually hinges on a correct specification of the distribution of the individual specific effect. Unfortunately, this is even more difficult in a dynamic model than in a static model.

To illustrate the main point, consider a dynamic version of (2) in which one of the explanatory variables is the lagged dependent variable y_{it-1} and the other explanatory variables are strictly exogenous. (2) would then specify the distribution of $(y_{i2}, ..., y_{iT_i})$ given the individual specific effect, the strictly exogenous variables and y_{i1}. However, it does not specify the distribution of y_{i1} given the individual specific effect and the strictly exogenous variables. There are then essentially two approaches. One approach, which was proposed by Heckman (1981), is to specify a separate model for y_{i1} given the individual specific effect and the strictly exogenous variables. A distributional assumption on the individual specific effect (conditional on the strictly exogenous variables) is therefore sufficient for one to proceed by maximum likelihood (or some other parametric method). The other approach, which was advocated by Wooldridge (2000), is to specify the distribution of the individual specific effect conditional on the strictly exogenous variables and on the first y, y_{i1}. With that, one can derive the distribution of $(y_{i2}, ..., y_{iT_i})$ given the strictly exogenous variables and

[11] In other words, it does not make sense to make assumptions of the distribution of $(\varepsilon_{is}, \varepsilon_{it})$ conditional on (x_{is}, x_{it}).

y_{i1}. The latter can then be used to estimate the parameters of interest (by maximum likelihood or some other parametric method).

Both the random effects and the fixed effects approach to dynamic nonlinear panel data models, have potential problems. But they also have some very appealing features. If (2) has been in effect before the start of the sample period, then the distribution of y_{i1} (given the random effect and the strictly exogenous variables) will depend on the joint distribution of the random effect and the strictly exogenous variables in periods prior to the start of the sample. It is almost unavoidable that modelling the distribution of y_{i1} (given the individual specific effect and the strictly exogenous variables) is inconsistent with (2) and one can at best hope that the approach will lead to a useful approximation.[12] On the other hand, there are cases in which there is a logical start of the process which coincides with the first time period in the sample. For example, the dependent variable might be the labor market status of high school graduates, and the data might contain information about the labor market status of indivuduals from the time they graduated from high school. In that case, there is no reason why one would want to specify the distribution of the first observation in a way that is consistent with (2) and the issues associated with this random effects approach are not different from those in a static model.

As pointed out in Wooldridge (2000), specifying the distribution of the random effect conditional on the strictly exogenous variables and on y_{i1}, can lead to very tractable functional forms for some common nonlinear models. On the other hand, in this setting the distribution of the random effect conditional on the strictly exogenous variables and on y_{i1}, is likely to be very complicated and to depend on all the exogenous variables in all time periods. The reason is that if the first observation depends on the random effect, then the distribution of the latter conditional on y_{i1} will typically depend on x_{i1}. Moreover, if past values of y_{it} are also generated from (2), then the distribution of the random effect conditional on y_{i1} and all values of x_{it} will depend (in a complicated way) on the values of x_{it} before the start of the sample. The distribution of the random effect conditional on the strictly exogenous variables and on y_{i1} will therefore depend on the time series properties of x_{it} is some very complicated way. In an attempt to overcome this, Arellano and Carrasco (1996) consider a model like (2) where the distribution of $\varepsilon_{it} + \alpha_i$ conditional on all the observables up

[12] Of course, that can be said about most almost any model econometric. The point here therefore is that the approach leads to one more level of approximation, and since the model for the first period is likely to be inconsistent with the model for the remaining periods, it might be difficult to interpret the results.

to time t is assumed to be homoskedastic normals (except that the variance may depend of t) but with unspecified expectation.

The difficulty of dealing with the initial conditions problem in random effects models makes it interesting to consider fixed effects models. Unfortunately, it turns out to be very difficult to make progress on these models in a fixed effects setting. Moreover, to the extend that progress has been made, the question remains whether estimating the parameters of the model allows one to calculate interesting marginal effects, since the latter will typically depend on the "structural" parameters, as well as the joint distribution of the individual specific effect and the initial observation.

To illustrate these issues consider a fixed effects dynamic logit model of the form

$$P\left(y_{it}|\,\alpha_i, x_i, y_{it-1}, y_{it-2}, ...\right) = \frac{\exp\left(x_{it}\beta + y_{i,t-1}\gamma + \alpha_i\right)}{1 + \exp\left(x_{it}\beta + y_{i,t-1}\gamma + \alpha_i\right)} \tag{15}$$

If there are at least four time periods, and the exogenous explanatory variables x_i are not present in (15), then Cox (1958) and Chamberlain (1985) have shown that one can estimate γ by considering the distribution of the data conditional on a sufficient statistic for α_i, which in this case is $\left(y_{i1}, y_{iT}, \sum_{t=1}^{T} y_{it}\right)$. Honoré and Kyriazidou (2000b) generalized this to the case where the logit model was also allowed to contain strictly exogenous explanatory variables.

As pointed out above, knowing γ and β in a model like (15) does not allow one to calculate the marginal effect of x_{it} on y_{it}. This is a limitation of the fixed effects approach. On the other hand, as discussed in Heckman (1978) and Arellano and Honoré (2001), there are cases in which $\gamma = 0$ is an interesting hypothesis. In those cases, it is interesting to estimate γ even if it does not allow one to calculate any marginal effects. Moreover, even though β does not allow one to calculate the magnitude of the effect of x_{it} on y_{it}, it does allow one to judge the relative importance of the different components of x_{it}, as well as to test whether the effect exists.

Honoré (1993) and Honoré and Hu (2000) consider censored panel data regressions models in which one of the dependent variables is a lagged dependent variable. Hu (2002) generalizes this to a censored panel data regression model in which one of the dependent variables is the lagged uncensored dependent variable. This model corresponds to a dynamic linear panel data model with top coding (in which it does not make sense to use the lagged censored dependent variable as an explanatory variable). Finally, Kyriazidou (2000) generalizes Kyriazidou (1997) by allowing both the selection equation and the outcome equation to depend on the lagged value of the dependent variable in the same equation.

4 Concluding Remarks

4.1 Computation

The adoption of new methods in econometrics is slowed down by the fixed cost of having to program the methods “from scratch”. Estimation of static random effects discrete choice and censored regression models is a canned command in Stata. As discussed in Wooldridge (2000), these routines can be used to also estimate dynamic models if one specifies the distribution of the individual specific effect conditional on the strictly exogenous variables and the first on y_{i1} in a particular manner. The fixed effects logit model can also be estimated in using a canned Stata command. The other estimators mentioned above require programming.[13]

4.2 Recommendations for Applied Work

For static models like the ones discussed in section 2, the choice between random effects and fixed effects models is similar to the choice between parametric and semiparametric models in cross sectional econometrics, and the pros and cons are also similar. Estimating a random effects panel data model or a parametric cross sectional model results in a fully specified model in which one can estimate all the quantities of interest, whereas fixed effects panel data models and semiparametric cross sectional models typically result in the estimation of some finite dimensional parameter from which one cannot calculate all functions of the distribution of the data.[14] Moreover, random effects models will usually lead to more efficient estimators of the parameters of the model if the distributional assumptions are satisfied. On the other hand, violation of the distributional assumptions in a random effects (or parametric) model will typically lead to inconsistent estimation of all the parameters. Fixed effects (and semiparametric) models make fewer such assumptions. Based on this it seems that if the main aim of an empirical exercise is to judge the relative importance of a number of variables or to statistically test whether certain variables are needed, and if efficiency is not too much of an issue, then a fixed effect approach is preferable because it will be less sensitive

[13]Some programs are available (typically in Gauss) for some of the estimators. See, for example, `http://www.econ.ucla.edu/kyria/` and `http://www.princeton.edu/~honore/pantob/`.

[14]This statement is somewhat misleading, because one could imagine estimating the “structural” paramaters, as well and the distributions of the unobservables in both fixed effects panel data models and semiparametric cross sectional models.

to distributional assumptions. On the other hand, if, as is often the case, one wants to use the model for prediction or for calculating the effect of various "what–if's", then a random effects model would be preferable. In that case, comparing the results to the ones obtained using a fixed effects approach can be used as a test (formal or informal) of the validity of the distributional assumptions made in the random effects model.

The comments above also apply to dynamic models of the type discussed in Section 3, except that both the random effects and fixed effects approaches have additional difficulties. The fixed effects approach suffers from a lack of knowledge about how to estimate the models, whereas the initial conditions problem is an additional issue for the random effects approach.

4.3 Open questions

There are numerous open questions in the literature on nonlinear panel data models. As already mentioned, many of the fixed effects methods do not lead to estimates of all the quantities that one needs to calculate the effect of x_{it} on the distribution of y_{it} holding everything else equal, whereas a random effects approach forces one to make distributional assumptions that in some situations may be undesirable. The recent paper by Altonji and Matzkin (2001) makes an important contribution by focussing directly on the effect of x_{it} on the distribution of y_{it} in a fixed effects model. Generalizing this to dynamic models would be an interesting topic for research.

4.4 Concluding Remark

Panel data methods are necessary for understanding individual dynamic behavior. Despite the difficulties associated with their use, they are likely to continue to play an important role, and it will be very valuable to expand the set of tools in this area. Panel data methods are also useful in situations that are cross sectional in nature. For example, Case, Lin, and McLanahan (2000) uses a fixed effects approach to control for the characteristics of the mother in a study of the educational attainment of children raised by step, adoptive or foster mothers compared to the birth children of the same women.

References

Ahn, H., and J. L. Powell (1993): "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58(1-2), 3–29.

Altonji, J. P., and R. L. Matzkin (2001): "Panel Data Estimators for Nonseperable Models with Endogenous Regressors," NBER Technical Working Paper 267.

Andersen, E. B. (1970): "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society*, Series B, 32, 283–301.

Arellano, M., and S. R. Bond (1988): "Dynamic Panel Data Estimation Using DPD-A Guide for Users," 88/15 Institute for Fiscal Studies, London.

Arellano, M., and R. Carrasco (1996): "Binary Choice Panel Data Models with Predetermined Variables," CEMFI, Madrid.

Arellano, M., and B. E. Honoré (2001): "Panel Data Models. Some Recent Developments," *Handbook of Econometrics*, 5, Elsevier Science, Amsterdam.

Blundell, R., R. Griffith, and F. Windmeijer (1997): "Individual Effects and Dynamic Count Data," University College London.

Browning, M. (1992): "Children and Household Economic Behaviour," *Journal of Economic Literature*, 30, 1434–1475.

Case, A., I.-F. Lin, and S. McLanahan (2000): "Educational Attainment in Blended Families," Forthcoming, Evolution and Human Behavior.

Chamberlain, G. (1984): "Panel Data, in Griliches, Z. and M.D. Intriligator (eds.)," *Handbook of Econometrics*, 2, Elsevier Science, Amsterdam.

Chamberlain, G. (1985): "Heterogeneity, Omitted Variable Bias, and Duration Dependence," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman, and B. Singer, no. 10 in Econometric Society Monographs series,, pp. 3–38. Cambridge University Press, Cambridge, New York and Sydney.

Chen, S. (1998): "Root-N Consistent Estimation of a Panel Data Sample selection Model," Hong Kong University of Science and Technology.

Cox, D. R. (1958): "The Regression Analysis of Binary Sequences," *Journal of the Royal Statistical Society. Series B*, 20, 215–242.

Hahn, J. (2001): "The Information Bound of a Dynamic Panel Logit Model with Fixed Effects," *Econometric Theory*, 17(5), 913–32.

Hajivassiliou, V. A., and P. A. Ruud (1994): "Classical Estimation Methods for LDV Models Using Simulation," in *Handbook of econometrics*, ed. by R. F. Engle, and D. L. McFadden, no. 4 in Handbooks in Economics,, pp. 2383–2441. Elsevier, North-Holland, Amsterdam, London and New York.

Hausman, J., B. H. Hall, and Z. Griliches (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, 52(4), 909–38.

Heckman, J. J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5(4), 475–92.

——— (1978): "Simple Statistical Models for Discrete Panel Data Developed and Applied to Tests of the Hypothesis of True State Dependence against the Hypothesis of Spurious State Dependence," *Annales de l'INSEE*, pp. 227–269.

——— (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–61.

——— (1981): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," *Structural Analysis of Discrete Panel Data with Econometric Applications, C. F. Manski and D. McFadden (eds)*, pp. 179–195.

Honoré, B. E. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533–565.

——— (1993): "Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependent Variables," *Journal of Econometrics*, 59, 35–61.

Honoré, B. E., and L. Hu (2000): "Estimation of Cross Sectional and Panel Data Censored Regression Models with Endogeneity," Princeton University.

Honoré, B. E., and E. Kyriazidou (2000a): "Estimation of Tobit-Type Models with Individual Specific Effects," *Econometric Reviews*, 19, 341–66.

——— (2000b): "Panel Data Discrete Choice Models with lagged Dependent Variables," *Econometrica*, 68, 839–874.

Hu, L. (2002): "Estimating a Censored Dynamic Panel Data Model with an Application to Earnings Dynamics," Northwestern University.

Kyriazidou, E. (1997): "Estimation of a Panel Data Sample Selection Model," *Econometrica*, 65, 1335–1364.

——— (2000): "Estimation of Dynamic Panel Data Sample Selection Models," University of Chicago.

Lancaster, T. (1997): "Orthogonal Parameters and Panel Data," Brown University.

Manski, C. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 357–62.

Manski, C. F. (1975): "The Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205–228.

——— (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 313–333.

Newey, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62(6), 1349–82.

Neyman, J., and E. L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32.

Powell, J. L. (1986): "Symmetrically Trimmed Least Squares Estimation for Tobit Models," *Econometrica*, 54(6), 1435–60.

——— (1987): "Semiparametric Estimation of Bivariate Latent Models," Working Paper no. 8704, Social Systems Research Institute, University of Wisconsin–Madison.

——— (1994): "Estimation of Semiparametric Models," in *Handbook of econometrics*, ed. by R. F. Engle, and D. L. McFadden, no. 4 in Handbooks in Economics,, pp. 2443–2521. Elsevier, North-Holland, Amsterdam, London and New York.

WOOLDRIDGE, J. M. (2000): "The Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," Michigan State University.